

FIELD REPLICATION OF CLASSWIDE PEER TUTORING

CHARLES R. GREENWOOD, GRANGER DINWIDDIE, VORIS BAILEY,
JUDITH J. CARTA, DON DORSEY, FRANK W. KOHLER,
CHRIS NELSON, DAVID ROTHOLZ, AND DAN SCHULTE

JUNIPER GARDENS CHILDREN'S PROJECT, BUREAU OF CHILD RESEARCH
UNIVERSITY OF KANSAS

We conducted a large-scale field replication study of classwide peer tutoring applied to spelling instruction (Greenwood, Delquadri, & Hall, 1984). Two hundred and eleven inner-city students in four schools participated during their first- and second-grade school years. The effects of classwide peer tutoring were compared to teacher instructional procedures and pretest probes using a group replication design (Barlow, Hayes, & Nelson, 1984). Analysis of group and individual results indicated that (a) both teacher instructional procedures and classwide peer tutoring were effective in increasing spelling performance above pretest levels, (b) peer tutoring produced statistically greater gains relative to the teachers' procedures for both low and high student groups formed on pretest levels, (c) these outcomes were representative of groups, classes, individuals, and years during the project, and (d) participant satisfaction with the program was generally high. A separate analysis of the social importance of treatment outcome revealed differential findings for low and high groups related to pretest levels. Implications of these findings are discussed.

DESCRIPTORS: field replication, spelling tutoring, longitudinal research, peer tutoring, fidelity issues

The use of tutoring methods in classroom instruction has been increasingly reported in the literature (e.g., Delquadri, Greenwood, Whorton, Carta, & Hall, 1986; Kalfus, 1984). Evaluations of tutoring have reported positive academic and social effects for tutees as well as for the tutors (e.g., Greer & Polirstok, 1982; Maheady & Harper, in press). Peer tutoring methods have been reported to be more effective than some traditional teaching methods (e.g., Greenwood, Dinwiddie, et al., 1984).

When compared with teacher instruction, peer tutoring increased rates of task presentations, tutee responses, and the immediacy of error correction (Greenwood, Dinwiddie, et al., 1984; Young, Hecimovic, & Salzberg, 1983), characteristics that may contribute to the positive academic and social effects of peer instruction. Recent studies have also reported that peer tutoring is more cost effective than other methods of improving student achievement, including computer-assisted instruction, reducing class size, or increasing learning time (Levin, Glass, & Meister, 1984).

Peer tutoring methods are characterized by peer pairing strategies and specific peer teaching repertoires. Peer tutoring has included cross-aged pairing (e.g., Willis, Crowder, & Morris, 1972), same-aged pairing (e.g., Dineen, Clark, & Risley, 1977), and classwide tutoring (e.g., Delquadri, Greenwood, Stretton, & Hall, 1983). Peer tutoring procedures have ranged from those left entirely to the discretion of the tutor (e.g., Niedermeyer, 1970) to systematic protocols including task presentation and error correction procedures that tutors have been trained to implement with high fidelity (e.g., Greenwood, Dinwiddie, et al., 1984).

This work was supported by grant No. 03144 from the National Institute of Child Health and Human Development. The authors thank principals L. Brown, R. Harrington, and N. Kibbey and teachers L. Grey, D. Harrison, P. Patterson, G. Dobson, P. Copeland, L. Gray, K. Zupan, D. Wilson, J. Rider, B. Klugh, M. Kline, S. Powell, I. L. Colgan, M. Robinson, and J. Wessels who participated in this project. Thanks are also due to Don Moritz, Director of Research in Kansas City Kansas School District #500. A debt is also owed Carmen Arreaga-Mayer, Barbara Terry, Hyman Hops, and Dennis Madrid for their comments on several drafts of this paper and the approach used in the analysis of treatment failures. The authors would also like to thank Reviewer D for pointing out the possibility of an analysis of treatment failures.

Reprints of this article can be obtained from Dr. Charles R. Greenwood at the Juniper Gardens Children's Project, 1614 Washington Blvd., Kansas City, Kansas 66102.

However, there have been few studies on either the long-term or large-scale use of peer tutoring (Gerber & Kauffman, 1981; Kalfus, 1984; Kauffman, Nussen, & McGee, 1977). A number of factors may have contributed to the absence of such research, including the failure of educational researchers to disseminate specific, effective procedures (Greer, 1982); the failure of researchers and developers to maintain the institutional support necessary for large-scale and longitudinal investigations; and the perception of teachers and educators that peer tutoring is only a temporary instructional arrangement and simply not a major instructional strategy.

A related methodological difficulty involves confounds that threaten the internal validity of longitudinal research, especially that conducted over periods longer than 1 school year (Campbell & Stanley, 1966; Greenwood, Hops, & Walker, 1977). These confounds are primarily due to the nonequivalence of conditions over time, for example, natural variations in treatment agents (i.e., changes in teachers), changes in the time devoted to instruction in specific content areas (e.g., spelling versus other subject matter), and variations in instructional procedures.

The purpose of this study was to examine the effects of a large-scale and long-term implementation of classwide peer tutoring procedures on spelling achievement and participant satisfaction. This study sought to extend prior research (e.g., Delquadri *et al.*, 1983; Greenwood, Dinwiddie, *et al.*, 1984) that was based upon relatively few students and teachers. A field replication design was used to establish the subject generality of tutoring effects across 211 students, four schools, 16 teachers, and 2 school years. The specific questions addressed were (a) was there a statistically significant difference in spelling performance produced by teacher versus classwide peer tutoring methods of instruction? (b) was there a statistically significant difference between the peer tutoring procedures used in both years? (c) was there a difference in the relative effectiveness of these procedures for low pretest versus high pretest students? (d) was it possible to isolate factors that may have accounted for

treatment failure or differential outcomes? (e) were the results generally representative of individual classes and students over years? and (f) were teachers and students satisfied with the classwide peer tutoring program?

METHOD

Subjects and Personnel

The students in this study were 211 inner-city, elementary-aged children who participated with parent permission during their first- and second-grade years. This represented 97% of all students in these classrooms during these years. For purposes of a group analysis, the students who participated in Year 2 ($n = 174$) were assigned to either high pretest or low pretest groups based upon their average performance on weekly spelling pretests given in Year 2. Students who scored a mean of 50% or more on these pretests were assigned to the high pretest group ($n = 104$, $M = 71.2\%$), those below 50% were assigned to the low pretest group ($n = 70$, $M = 29.3\%$).

Sixteen female teachers participated, eight during first grade and eight during the second grade. Six consultants trained the teachers to implement the classwide peer tutoring program and monitored their classroom implementation during each year.

Setting

The study took place in four inner-city schools that served a low socioeconomic, largely black population in the northeast section of Kansas City, Kansas. All four schools qualified for Chapter I status and received federal funds for free lunches and support of resource programs in reading and mathematics instruction. The percentage of students receiving Chapter I services in each school were 50%, 44%, 45%, and 47%, respectively. The study took place in regular classrooms during daily 30- to 45-min spelling periods.

Assessment

Weekly spelling tests. A standard spelling posttest was administered every Friday to assess students' mastery of the spelling words taught each

week. The teachers administered these 10- to 20-item spelling tests by pronouncing each word, using it in a sentence, then pronouncing it a second time. Students wrote the words on writing paper. The tests were corrected either by the teacher or by the students at the teacher's discretion. Individual students' percent correct scores on the posttest were publicly posted on a chart in each classroom.

During Year 2, teachers also administered a pretest on alternate Mondays covering the new words to be taught that week. This provided a check on the level of difficulty of words selected by teachers and a basis for demonstrating posttest gains. The results of the Monday tests were not publicly posted and no contingencies were established based on these data.

Reliability of tests. Randomly selected tests (54 in Year 1, 96 in Year 2) were regraded and rescored by program consultants. Pearson r correlations were computed between teacher or student percent correct scores on individual tests and those obtained by program consultants. The correlation was 0.88 for the percent correct in Year 1 and 0.97 in Year 2. A correlated t -test comparing teacher/student and consultant scores across both years was not statistically significant, $t(148) = 0.21$, $p > .05$. This indicated that teachers or students and consultants produced percent correct scores of equal magnitude on the tests.

Teacher procedural checklist. A 36-item checklist was developed to assess the fidelity of each teacher's implementation of the peer tutoring program (Carta, Dinwiddie, Kohler, Delquadri, & Greenwood, 1984). The checklist was organized into three sections including items assessing (a) the presence of specific program materials (e.g., posted point charts), (b) the correct sequence and occurrence of teacher behaviors (e.g., setting a timer, telling students to begin tutoring, giving bonus points), and (c) correct peer tutoring behaviors (e.g., presenting the spelling word, implementing point contingencies, correcting errors). During direct observations of tutoring by consultants, items on the checklist were scored as present or absent. A total score for each of the three sections (i.e., materials, teacher procedures, and student procedures) and a

total composite percentage score were obtained. The first check was completed 1 week after the teacher had completed training and initiated the program. Subsequent checks occurred at the consultant's discretion.

Teacher satisfaction survey. A 20-item survey was completed by teachers at the end of each school year. These items covered general satisfaction with the entire program (e.g., "The tutoring program was helpful for students at all ability levels in my class") and specific items related to program components that might be improved (e.g., "Students were able to move to their partners quickly and quietly"). Teachers recorded their opinions using a 5-point Likert scale with the following values: 1 = strongly disagree, 2 = disagree, 3 = not sure, 4 = agree, and 5 = strongly agree.

Student satisfaction. A 14-item survey was administered to students at the end of each school year. The survey contained seven items dealing with general program satisfaction and seven items dealing with peer relations. All of the questions in the survey were in the form "How did you like . . . ?" For example, two items were "How did you like tutoring this year?" and "How did you like working with a partner?"

Each item was evaluated using a Likert scale consisting of three rating choices: 1 = frowning face, 2 = neutral face, and 3 = smiling face. The survey was administered by the teacher who read the instructions, demonstrated marking of the student's response sheet, and had the students complete two practice items before completing the survey.

General Procedures

In order to conduct a longitudinal project, it was necessary to ensure that teachers in both first and second grades were willing to participate. For the second-grade teachers, this meant waiting a year after agreeing to participate before receiving training and implementing the program. This agreement was obtained for all teachers prior to the start of the study. However, two teachers, one in each year, refused to participate in subsequent meetings with project consultants. This accounts for the large dif-

ferences in the number of participants between Year 1 and Year 2 at School 1 and School 4 (see Table 1).

At the end of the first-grade year, consultants met jointly with the participating first- and second-grade teachers in order to review progress made by the first-grade teachers. At this time, a schedule for training the second-grade teachers in the following year was established.

Design and Specific Procedures

A field replication design with single-subject design features (i.e., baseline, reversal, and pretest probes) was used (Barlow *et al.*, 1984; Kazdin, 1982). However, the specific designs used in each class and across individual students over the 2 years varied. For example, the number of baseline-reversal phases received by individual students ranged from one to four. Since the emphasis of the study was the long-term implementation of the classwide peer tutoring procedures, baseline and reversal probes were kept to a minimum duration, ranging from 1 to 3 weeks per phase.

Teacher procedures (A). The teacher procedures phase was an assessment of student spelling performance when taught using the teacher's regular procedures. Tutoring was not used. Teachers' procedures varied but did contain several common elements, including (a) use of a spelling text with word lists and vocabulary exercises to be completed, (b) chalkboard or overhead transparency discussions of new word lists, (c) self study, (d) oral spelling with teacher assistance, and (e) home assignments to study the words.

Classwide peer tutoring (B). Prior to the classwide peer tutoring phase, teachers read a manual (Carta *et al.*, 1984) that described the implementation of classwide peer tutoring and discussed its implications with the consultant.

Students were taught the tutoring procedures during three class periods, totaling 60 min. The consultant and teacher modeled correct tutoring procedures and gave the students feedback on their attempts to imitate the procedures. Bonus points and praise were delivered contingent on correct tutoring during ongoing spelling instruction. In Year

2, student training required much less time because all but students new to the school that year had previously used tutoring. In Year 2, the procedures were orally reviewed in one session and correction with feedback was provided during the actual tutoring sessions.

The general tutoring procedures involved (a) weekly lists of 10 (Year 1) or 20 (Year 2) new spelling words, (b) new partners each week, (c) random pairing of partners, (d) immediate error correction, (e) contingent individual point earning, (f) two teams competing for the highest cumulative point total, (g) winning team social reward, and (h) public posting of individual and team total scores (Delquadri *et al.*, 1986).

Tutoring required 30 min to implement each day. Each student served as a tutor for 10 min, followed by 10 min as the tutee. Five to 10 min were needed to add and post individual and team points. Students were randomly assigned to one of two teams every Monday, and remained on these teams for the entire week. Weekly restructuring of teams assured that all children were intermittently on a winning team.

After a signal from the teacher, each tutor presented the first word to be spelled (e.g., the tutor said "Spell cat"). The tutee responded by simultaneously spelling the word aloud and writing the word. The tutor awarded two points for spelling and writing a word without error. If an error occurred, the tutor spelled the word correctly. If the tutee then wrote the word correctly three times, the tutor awarded one point. If an additional error was made by the tutee at this step, no points were earned and a new word was presented by the tutor.

The teacher moved among the students providing assistance and awarding bonus points to tutors contingent on correct tutoring behaviors and to tutees for responding immediately and for working cooperatively with their tutor. This continued until the end of the 10 min, when rules were reversed. A second tutoring period began and followed the same procedures. The teacher recorded students' points on a large chart and calculated team totals. The winning team was applauded as was the losing team for making a good effort. On rare occasions,

teachers used nonsocial backup consequences for winning teams.

Data Analysis

The weekly spelling test scores for individual subjects were reduced to condition means and statistical summaries completed. These condition means were teacher procedures (Tpro1) and classwide peer tutoring (Cwpt1) in Year 1; and pretest (Pre2), teacher procedures (Tpro2), and classwide peer tutoring (Cwpt2) in Year 2. These data were then summarized as means for teachers by years and for high and low pretest groups.

Because the number of students varied across years as did the conditions implemented by teachers, a cross-sectional statistical analysis of the entire sample was not considered appropriate.¹ (In order to conduct the statistical analysis with as large a sample as possible and to avoid missing data, the teacher procedure spelling performance was consolidated over both years. Thus, this mean score may reflect Year 1 performance, Year 2 performance, or the average performance based on both years for any specific student.) By combining the teacher procedure phases across years (Tpro1 and 2), it was possible to create a longitudinal analysis for 99 students in the program for both years with complete data. A two-factor univariate analysis of variance (ANOVA) was computed to examine differences between one repeated measures factor, *Instructional Procedures* (i.e., Tpro1 and 2, Cwpt1 and 2) and one nonrepeated factor, *Groups* (i.e., high pretest versus low pretest). When either main effect or interaction effect differences were significant, pairwise comparisons among means were examined using the Tukey HSD test at the .01 level (Kirk, 1968). (The Tukey critical range for comparisons in means at $\alpha = .01$ was 4.35.)

In order to examine the social importance of these effects for students, the ANOVA for repeated measures was followed by an analysis of treatment success and failure (see Barlow et al., 1984). Success

and failure in the present investigation were defined as gains in spelling performance under the tutoring procedure relative to gains under the teacher procedure. For this purpose, an index of improvement based on mean performance under Pre2, Tpro1 and 2, and Cwpt2 conditions was used and a cutpoint developed for dividing students into two groups (treatment failures and successes) based on their relative performance across these instructional conditions. The index was defined by $[(Cwpt2 - Pre2) - (Tpro1 \text{ and } 2 - Pre2)]$.

On a 20-word test in Year 2, the cutpoint represented a minimum change equal to or less than 10%, or at least two words per test. Thus, a student was considered a treatment failure if he or she failed to reach this magnitude of gain under classwide tutoring. Alternately, a student was considered a treatment success if he or she performed at or higher than this level during tutoring.

A chi-square was performed to test the hypothesis that differential numbers of students were either treatment successes or failures in the high versus low pretest groups. This test was followed by additional exploratory two-way ANOVA tests, between groups and treatment outcome (i.e., success versus failure), on several additional variables in an attempt to isolate potential explanations for these outcomes.

RESULTS

Fidelity of Program Implementation

In Year 1, the teachers implemented a mean of 76.5% of the prescribed peer tutoring components (range, 64% to 92%). Subsequent checks averaged 82.7% across teachers. Two teachers were retrained by consultants in Year 1 and additional training focusing on student compliance during activity transitions was provided. In one instance, the teacher modified the procedure by eliminating competing teams because of ethical concerns. Thus, in this classroom (Teacher 1 in Year 1), the class sought to equal or exceed their point total from day to day as a single team rather than to compete as two rival teams.

In Year 2, the equivalent figure on the first

¹ These analyses would require several independent tests each with a separate error term. This procedure would increase the overall risk of experiment-wide error, i.e., some differences would be significant by chance.

fidelity check following training was 82.8%, ranging from 57% to 97%. The average was 90.6% for teachers on all subsequent checks. No teachers in Year 2 received additional training or retraining.

Spelling Achievement

For the 99 students with complete data, the mean percentage of correct responses under each procedure was 74.9% (Tpro1 and 2), 85.0% (Cwpt1), and 89.0% (Cwpt2). The ANOVA yielded a significant main effect for *Instructional Procedures*, $F(2, 194) = 37.1, p < .001$. Pairwise post-hoc comparisons of the *Instructional Procedures* means indicated that during tutoring, students in both years made statistically significant gains ($p < .01$) in spelling performance relative to the teacher procedures. There was no difference between tutoring in Years 1 and 2 as students maintained high levels of performance in both years.

The main effect for *Groups* produced means of 75% for the low pretest group versus 90% for the high pretest group, $F(1, 97) = 31.9, p < .001$. Thus, the initial difference of 40% between groups at Pre2, 30% versus 70%, $F(1, 95) = 200.4, p < .001$, was substantially reduced but not entirely removed as a result of both teacher and tutoring procedures. Although the *Instructional Procedures* by *Groups* interaction effect was not significant, it was interesting to note that this difference between group means progressively shrank over procedures and years; Tpro1 and 2 (difference = 20.2%), Cwpt1 (difference = 14.9%), and Cwpt2 (difference = 13.9%).

The effects for all students by teacher and year are illustrated in Table 1. Like the statistical analysis, these data indicate that the highest spelling accuracy levels were obtained during the tutoring program. The mean percentages for all students across conditions were as follows: Year 1, 78.1 (Tpro1) versus 84.1 (Cwpt1), and 54.8 (Pre2) versus 70.0 (Tpro2) versus 88.5 (Cwpt2) in Year 2. The tutoring procedures were also associated with higher performance than the teacher procedures for data grouped according to low and high pretest levels and for individual classrooms in Years 1 and 2.

To illustrate experimental control at the indi-

vidual student level, the data from two randomly selected students are presented in Figure 1. These students reflected the range in single-subject designs used in the study. Student 3 received four condition changes, representing the maximum number of experimental manipulations, while Student 8 experienced one condition change representing the minimum number of manipulations. These data document higher spelling scores under classwide peer tutoring than under either pretest or teacher procedure conditions.

Participant Satisfaction

The teacher rating means over six general satisfaction items ranged from 3.9 to 4.6 in Year 1 and from 3.3 to 4.3 in Year 2 on a 5-point scale. These results reflected generally positive evaluations of the peer tutoring program.

Most items on the student satisfaction survey received a rating of 2.4 or better on a 3-point scale, also indicating general satisfaction with the program and peer relations during tutoring. A notable exception to these generally positive evaluations was the rating of the error correction procedure used by tutors, which received a mean of 2.0 in both years.

Analysis of Treatment Success and Failure

Of the 99 students with complete data, 53% were classified as treatment failures (less than 10% improvement under peer tutoring relative to teacher procedures), whereas 47% of students improved by 10% or more under peer tutoring. A differential pattern of improvement was noted for high and low pretest groups ($\chi^2 = 5.2, p < .02$). In the high group, 35 (63%) were classified as treatment failures and 21 (37%) as treatment successes. Equivalent values for the low group students were 17 (39%) failures and 26 (61%) successes. Thus, a greater number of treatment successes occurred for low pretest students (61%) than for high pretest students (37%).

Factors that differentiated between treatment successes and treatment failures proved difficult to isolate. There were no statistical differences between the groups in days absent or satisfaction with the tutoring program. Interestingly, the treatment failure group had significantly higher pretest scores

Table 1
Phase Mean Summary for All Students by Teacher, School, Year, and Condition

Teacher	School	Year	Teacher procedure 1			Classwide peer tutoring 1			Pretest 2			Teacher procedure 2			Classwide peer tutoring 2		
			M	SD	n	M	SD	n	M	SD	n	M	SD	n	M	SD	n
1	1	1	—	—	—	88.3	12.1	8									
2	1	1	—	—	—	74.2	22.9	28									
3	1	2							69.1	19.2	24	47.8	20.2	24	86.2	14.1	24
4	1	2							56.6	25.5	22	80.0	21.9	22	87.8	14.2	22
5	1	2							44.3	20.2	23	81.0	19.5	22	86.3	16.2	23
6	2	1	83.3	21.0	23	93.0	12.4	23									
7	2	1	70.4	25.7	22	81.6	16.9	22									
8	2	2							48.3	25.3	14	81.0	16.1	14	90.6	8.8	14
9	2	2							60.2	20.9	21	80.3	27.6	21	87.9	17.1	21
10	3	1	55.8	30.9	23	76.5	22.7	23									
11	3	1	78.1	23.9	21	84.4	16.6	21									
12	3	2							49.1	28.2	24	69.1	27.4	26	86.8	15.0	24
13	3	2							52.5	26.3	24	41.9	28.7	23	91.3	9.8	24
14	4	1	90.2	15.1	23	92.1	16.7	23									
15	4	1	81.1	25.8	21	87.6	15.5	21									
16	4	2							56.0	23.7	22	—	—	—	91.7	16.9	22
High pretest			89.3	18.2	80	91.2	13.6	102	71.2	13.4	104	78.6	25.1	91	94.7	6.6	104
Low pretest			64.9	28.1	53	75.1	21.4	67	29.3	13.3	70	58.7	28.3	61	79.1	17.4	70
Overall			78.1	26.5	133	84.1	18.9	169	54.8	24.5	174	70.0	28.3	152	88.5	17.3	174

Note. — = Data were missing due to the design used in this classroom.

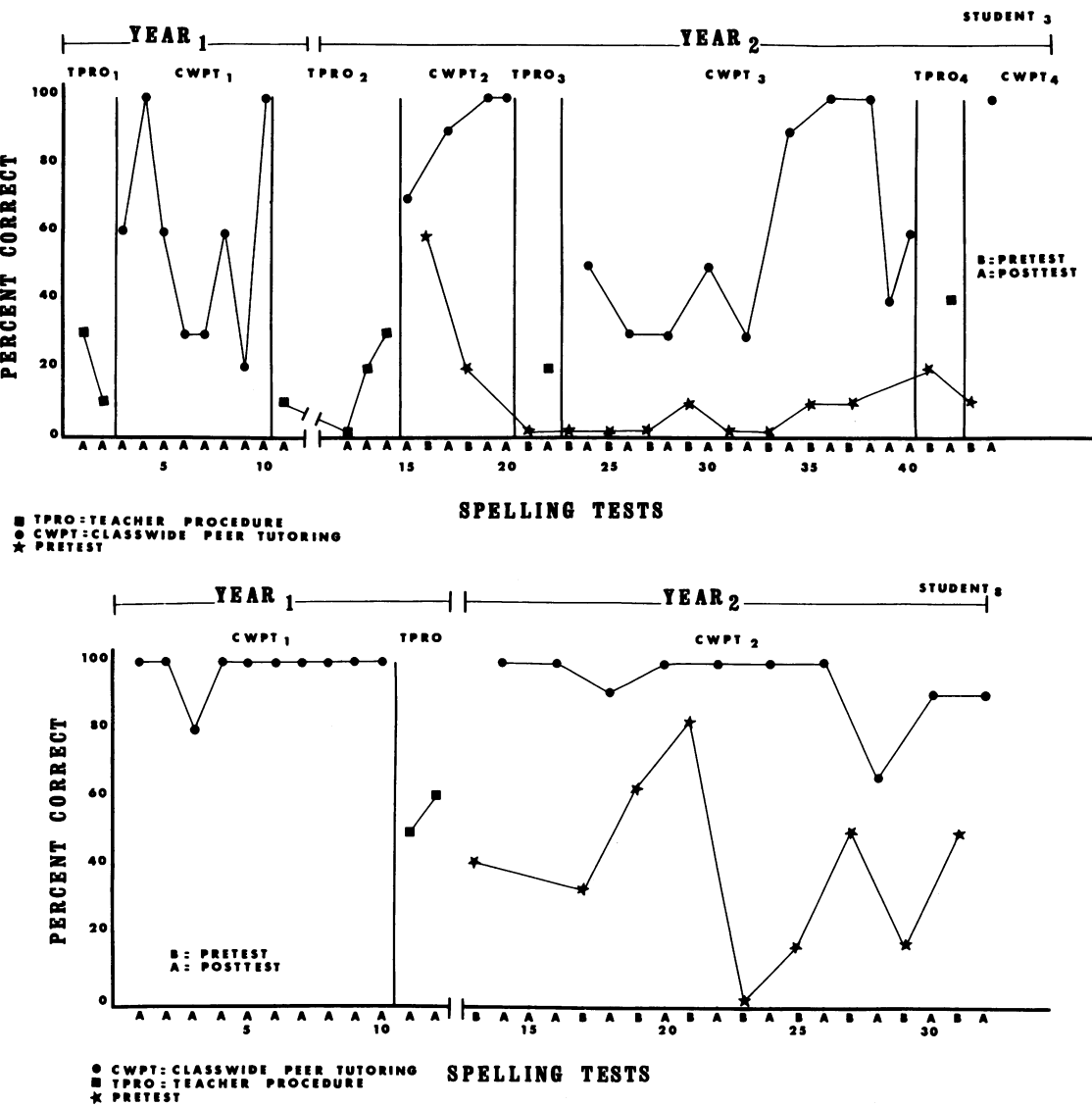


Figure 1. The percentage of correctly spelled words on teacher-prepared spelling tests over 2 school years for Student 3 (upper panel) and Student 8 (lower panel). These students were randomly selected to illustrate the range of single-subject design and subject effects across the entire study.

($M = 60.9\%$) than the treatment success group ($M = 45.2\%$), $F(1, 95) = 5.6$, $p = .02$. This suggested that ceiling effects may have precluded sufficient performance improvements to qualify as a treatment success for some students.

DISCUSSION

The purpose of this investigation was to summarize the effects of a large-scale and long-term

field replication of classwide peer tutoring on spelling achievement and participant satisfaction over 2 school years. Results indicated that substantial improvements were made by students over pretest levels under teacher instructional procedures, but that significant additional improvements were made during classwide peer tutoring. These improvements were made by both high and low student groups and were general across classes, individuals, and years. Teacher and student satisfaction with

the program was noted to range from moderate to high levels. These results tended to replicate prior studies using smaller samples and implemented for shorter periods of time (e.g., Greenwood, Dinwiddie, et al., 1984).

It should be noted, however, that these conclusions are based on a field replication approach that includes a rather diverse set of individual data and differences in experimental design and the timing of pretest assessments between classrooms. Because pretest data were available only in Year 2 and were not systematically probed for all students during the brief teacher procedure phases that year, caution is advised when interpreting the magnitude of gains as they may not be exact. This is because difficulty of the spelling words was not controlled throughout the entire study.

Analysis of treatment success and failure yielded several points. First, a principal reason for the lower numbers of treatment successes in the high pretest group was due to a ceiling effect. Of this group, 12% had pretest scores of 90% or greater and thus could not achieve gains large enough to be classified as a treatment success. Furthermore, the treatment failure group was comprised of more high pretest students (60.9%) than the treatment success group (45.2%). Efforts to individualize the difficulty of spelling content could solve this problem in subsequent research. Second, the data on student absences and satisfaction with tutoring were not statistically related to either pretest groups or levels of outcome. Third, the data on fidelity of tutoring implementation, although documenting a high-quality implementation of the program, were not sensitive to the tutoring interactions of individual students. The possibility that variations in the use of tutoring procedures might relate to outcome differences requires further research.

Collectively, the results of this large-scale and long-term project demonstrated both statistically and practically that better performance was obtained within a fixed time period with the use of effective instructional procedures (Skinner, 1984). This is the type of demonstration that stands in contrast to recent arguments produced by blue-ribbon education panels calling for increasing the

length of the school year rather than improving teaching procedures (e.g., National Commission on Excellence in Education, 1983).

REFERENCES

- Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). Clinical replication. In D. H. Barlow, S. C. Hayes, & R. O. Nelson (Eds.), *The scientist practitioner: Research and accountability in clinical and educational settings* (pp. 290-346). New York: Pergamon.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Carta, J. J., Dinwiddie, G., Kohler, F., Delquadri, J., & Greenwood, C. R. (1984). *The Juniper Gardens class-wide peer tutoring programs for spelling, reading, and math: Teacher's manual*. Kansas City, KS: Bureau of Child Research, University of Kansas.
- Delquadri, J., Greenwood, C. R., Stretton, K., & Hall, R. V. (1983). The peer tutoring spelling game: A classroom procedure for increasing opportunity to respond and spelling performance. *Education and Treatment of Children*, 6, 225-239.
- Delquadri, J., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional Children*, 52, 535-542.
- Dineen, J. P., Clark, H. B., & Risley, T. R. (1977). Peer tutoring among elementary students: Educational benefits to the tutor. *Journal of Applied Behavior Analysis*, 10, 231-238.
- Gerber, M., & Kauffman, J. M. (1981). Peer tutoring in academic settings. In P. Strain (Ed.), *The utilization of classroom peers as behavior change agents* (pp. 155-188). New York: Plenum.
- Greenwood, C. R., Delquadri, J., & Hall, R. V. (1984). Opportunity to respond and student academic performance. In W. Heward, T. Heron, D. Hill, & J. Trap-Porter (Eds.), *Behavior analysis in education* (pp. 58-88). Columbus, OH: Charles E. Merrill Publishing Co.
- Greenwood, C. R., Dinwiddie, G., Terry, B., Wade, L., Stanley, S., Thibadeau, S., & Delquadri, J. (1984). Teacher- versus peer-mediated instruction: An ecobehavioral analysis of achievement outcomes. *Journal of Applied Behavior Analysis*, 17, 521-538.
- Greenwood, C. R., Hops, H., & Walker, H. M. (1977). The durability of student behavior change: A comparative analysis at follow-up. *Behavior Therapy*, 8, 631-638.
- Greer, R. D. (1982). Countercontrols for the American Educational Research Association. *The Behavior Analyst*, 5, 65-76.
- Greer, R. D., & Polirstok, S. R. (1982). Collateral gains and short-term maintenance in reading and on-task responses by some inner-city adolescents as a function of their use of social reinforcement while tutoring. *Journal of Applied Behavior Analysis*, 15, 123-139.

- Kalfus, G. R. (1984). Peer mediated intervention: A critical review. *Child and Family Behavior Therapy*, **6**, 17-43.
- Kauffman, J. M., Nussen, J. L., & McGee, C. S. (1977). Follow-up in classroom behavior modification: Survey and discussion. *Journal of School Psychology*, **15**, 343-348.
- Kazdin, A. E. (1982). *Single case research design: Methods for clinical and applied settings*. New York: Oxford University Press.
- Kirk, R. E. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Wadsworth.
- Levin, H., Glass, G., & Meister, G. (1984). *Cost-effectiveness of four educational interventions (Report No. 84-A11)*. Institute for Research in Educational Finance and Governance (IFG), Stanford University, Stanford, CA.
- Maheady, L., & Harper, G. (in press). A class-wide peer tutoring program to improve the spelling test performance of low-income, third and fourth grade students. *Education and Treatment of Children*.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- Niedermeyer, F. C. (1970). Effects of training on the instructional behaviors of student tutors. *Journal of Educational Research*, **64**, 119-123.
- Skinner, B. F. (1984). The shame of American education. *American Psychologist*, **39**, 947-954.
- Willis, J., Crowder, J., & Morris, B. (1972). A behavioral approach to remedial reading using students as behavioral engineers. In G. Semb (Ed.), *Behavior analysis in education* (pp. 211-221). Lawrence, KS: The University of Kansas Support and Development Center for Follow Through.
- Young, C. C., Hecimovic, A., & Salzberg, C. L. (1983). Tutor-tutee behavior of disadvantaged kindergarten children during peer teaching. *Education and Treatment of Children*, **6**, 123-135.

Received May 5, 1986

Initial editorial decision July 10, 1986

Revisions received October 10, 1986; December 12, 1986

Final acceptance January 6, 1987

Action Editor, R. Wayne Fuqua